

Figure 1

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3.5	0.121	0.036	0.467	0.664
4	0.096	0.06	0.519	0.708
4.5	0.078	0.079	0.565	0.745
5	0.062	0.098	0.615	0.782
5.5	0.05	0.127	0.659	0.813
6	0.04	0.163	0.694	0.836
6.5	0.033	0.202	0.725	0.855
7	0.025	0.248	0.763	0.878
7.5	0.021	0.304	0.78	0.889
8	0.015	0.368	0.816	0.909
8.5	0.012	0.418	0.836	0.92
9	0.009	0.512	0.856	0.93
9.5	0.007	0.581	0.863	0.934
10	0.006	0.679	0.835	0.919

Figure 2

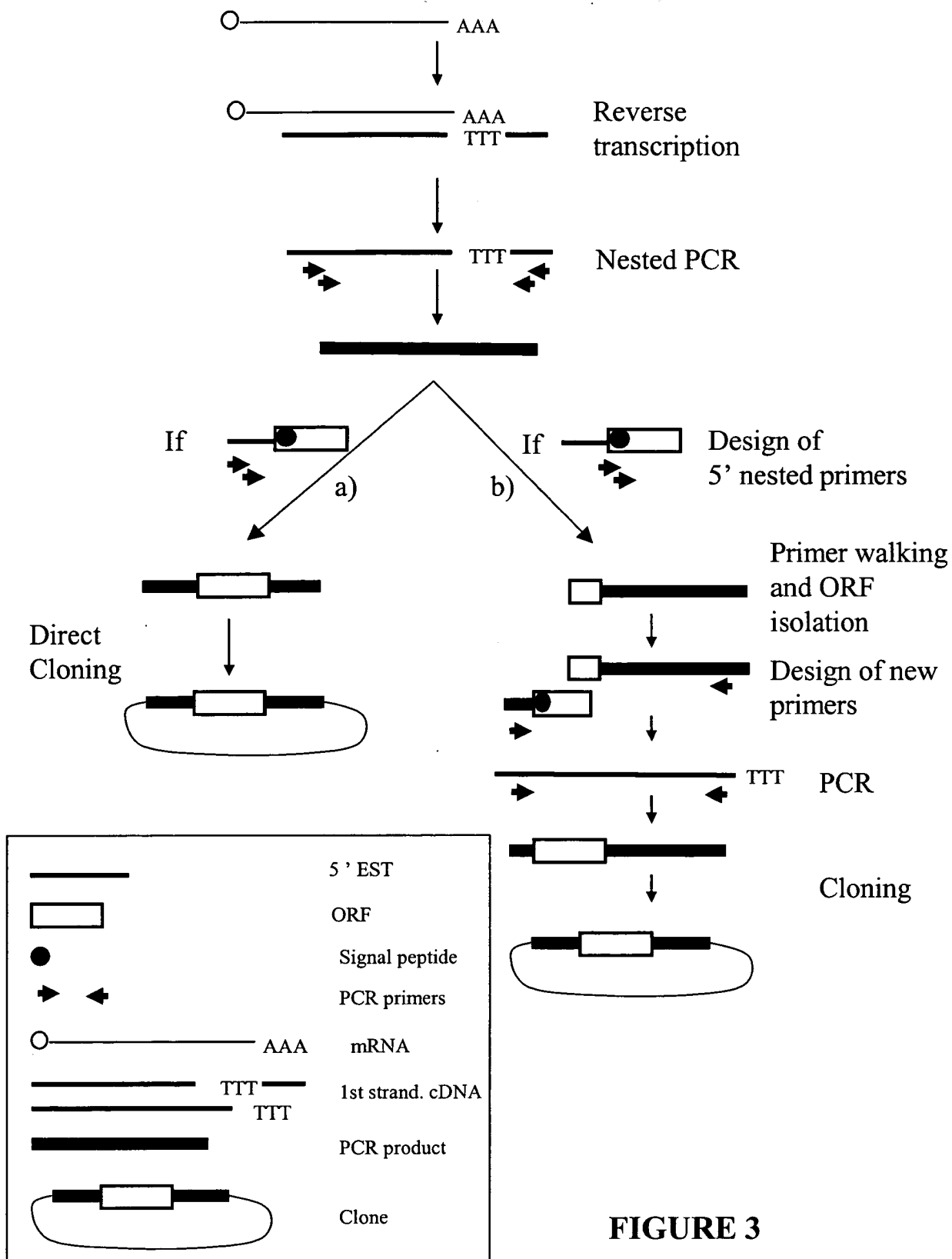


FIGURE 3

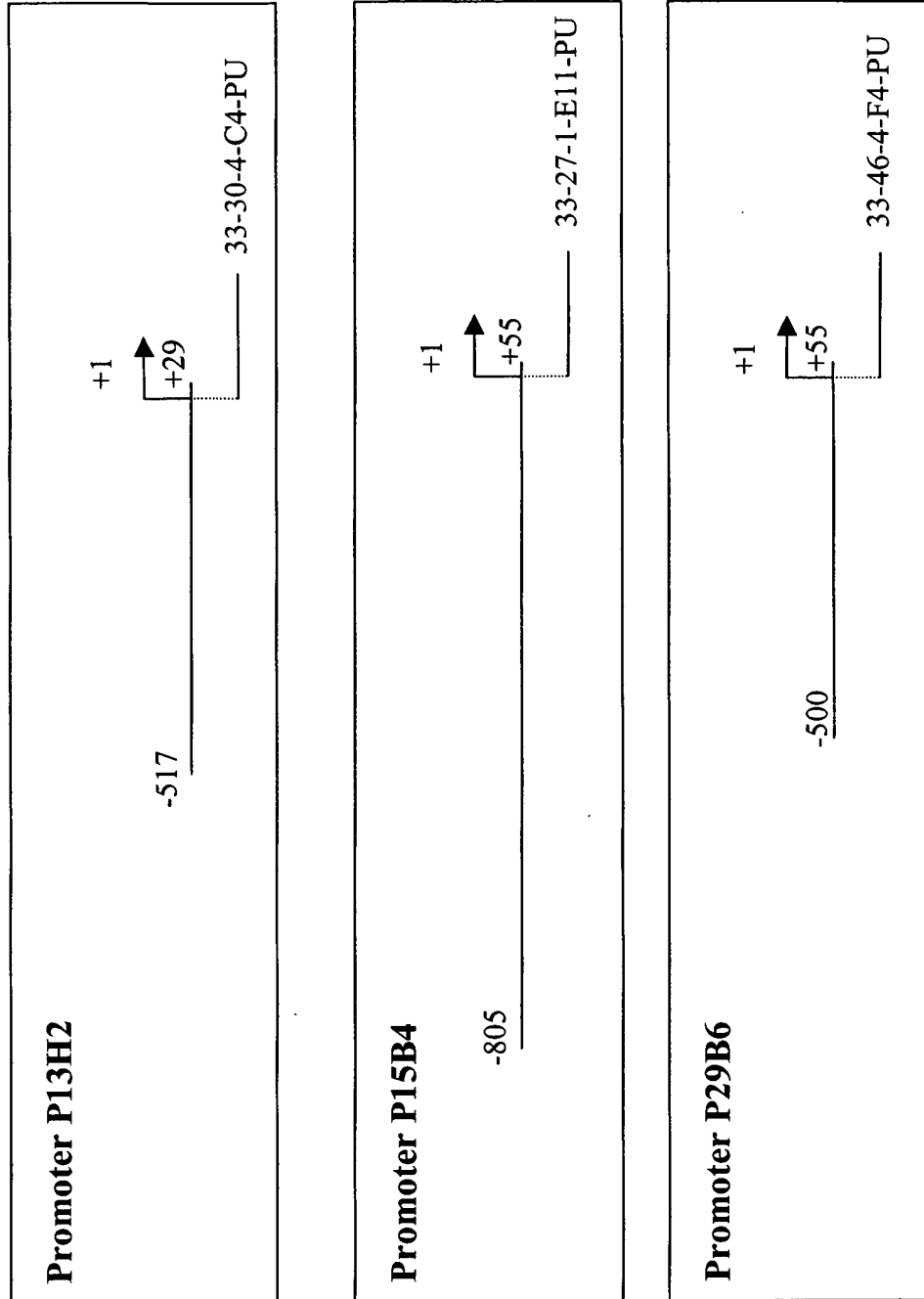


Figure 4

Search characteristic		Selection Characteristics		
Step	Program	Strand	Parameters	Identity (%) Length (bp) Comments
miscellaneous	FASTA	both	-	90 15
tRNA	FASTA	both	-	80 60
rRNA	BLASTN	both	S=108	80 40
mtRNA	BLASTN	both	S=108	80 40
Prokaryotic	BLASTN	both	S=144	90 40
Fungal	BLASTN	both	S=144	90 40
Alu	BLASTN	both	S=72, B=5	70 40 max 5 matches, masking
L1	BLASTN	both	S=72, B=5	70 40 max 5 matches, masking
Repeats	BLASTN	both	S=72	70 40 masking
PolyA	BLAST2N	top	W=6, S=10, E=1000, N=-12	90 10 in the last 100 nucleotides
Polyadenylation signal	-	top	AATAAA allowing 1 mismatch	in the 50 nucleotides preceding the 5' end of the polyA
Vertebrate	BLASTN then FASTA	both	-	90 then 70 30 first BLASTN and then FASTA on matching sequences
ESTs	BLAST2N	both	-	90 30
Geneseq	BLASTN	both	W=8, B=10	90 30
ORF	BLASTP	top	W=8, B=10	- on ORF proteins, max 10 matches
Proteins	BLASTX	top	E = 0.001	70 30

Parameters used for each step of cDNA analysis

Figure 5